DOCUMENT RESUME

ED 178 570 TM 009 715

AUTHOR Ahmann, J. Stanley
TITLE A Speculative View of the Needs in Educational Measurement in Higher Education in the 1980s.
PUB DATE [79]
NOTE 11p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; College Bound Students; Communication Problems; Criterion Referenced Tests; Educational Assessment; Educational Needs; Effective Teaching; *Evaluation Needs; *Futures (of Society); Higher Education; *Measurement Goals; *Needs Assessment; Norm Referenced Tests; Occupational Tests; Program Evaluation; *Student Testing; Test Interpretation; Test Results; Undergraduate Study

ABSTRACT

The focus of educational measurement in higher education in the 1980's is discussed: (1) better evaluation of student achievement; (2) improved communication of measurement results to various audiences; (3) evaluation of undergraduate educational programs; and (4) systematic assessment of the educational needs of college-bound students. In addition to the continued use of norm referenced tests, it is suggested that additional criterion referenced tests be developed, and that instructional goals and related subject matter be more closely defined. It is also recommended that item banks composed of thousands of test items be developed. Narrative reporting of test results is suggested as a means of improving the communication gap between test developers and faculty, students, and parents. Areas of evaluation of undergraduate programs are discussed: achievement measures of specific programs; occupational competency of students in professional programs, using written, oral, and performance tests; and evaluation of teaching effectiveness, rather than of teacher performance, partially based on student levels of achievement. To assess the needs of college-bound students, it is suggested that they be profiled as a distinct group, in a program similar to the National Assessment of Educational Progress. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. J. Ahmann

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED178570

TM009 715

A SPECULATIVE VIEW OF THE NEEDS IN EDUCATIONAL MEASUREMENT IN HIGHER EDUCATION IN THE 1980s

J. Stanley Ahmann
Iowa State University

The needs of educational measurement in higher education in the 1980s are massive. This fact alone can cause one to anticipate with at least modest excitement the developments in that decade that will inevitably occur. On the other hand, the size, number and importance of these needs is such that it is easy to become pessimistic about the likelihood that sufficient progress will indeed occur. It is probable, however, that the positive point of view is more justified than the negative one.

It is convenient to classify the major needs under four headings, namely:

1. The evaluation of student achievement
2. Communication of measurement results to multiple audiences
3. Evaluation of undergraduate educational programs
4. Systematic assessment of the educational needs of college-bound students

The Evaluation of Student Achievement

Although criterion-referenced achievement tests have been introduced and are used somewhat widely in elementary and secondary classrooms, the impact of this type of testing in the college classrooms has not been great. In the 1980s its use should increase and, as a consequence, greater interpretability of achievement results will occur.

Defining the learning domain on the basis of which criterion-referenced

tests are to be developed in higher education will be an even more difficult than it has proven to be in elementary and secondary schools. To develop successful criterion-referenced tests, faculty members will find it necessary to specify more clearly the goals of their instruction and the subject matter subdivisions related to those goals. A significant level of faculty training in this process will be required.

As is the case for all criterion-referenced tests, definition of mastery levels will be required. The question must be faced as to whether systematic means of accomplishing this can be developed, or whether--as stated by Glass (1978)--the most easily interpreted manifestation of student progress is change in levels of achievement.

It must be anticipated that the measurement of student achievement in higher education will still be dominated by norm-referenced tests in the 1980s. Given this, one must hope for more and better criterion-referenced interpretations of norm-referenced test data. Such interpretations are becoming comparatively common for elementary and secondary standardized achievement tests, and it is necessary that this practice be used increasingly in higher education.

Combining criterion-referenced tests and norm-referenced tests into a composite student achievement testing program is a proposal that has been encouraged, but rarely implemented, for some years. If it is believed that mastery of certain elements of an instructional program are needed and at the same time the "ceiling" for learning is very high, then combining mastery testing with developmental testing is a reasonable approach and should be used more frequently. In effect, this may mean selective use of criterion-referenced tests to determine whether students have

mastered basic elements of a college course, and norm-referenced tests in order to determine the ceiling level of achievement of various students (Trow, 1968).

Another feature of achievement testing that is underdeveloped in higher education is the preparation and use of test item banks (Shoemaker, 1975). Given the long-term testing needs in a content area, it is reasonable that continuous efforts be made to develop item-banks composed of not hundreds but possibly thousands of test items, from which computers can select a representative subset, which then becomes an achievement test to be administered to a given group of students for a stated purpose. This procedure permits a much greater flexibility to be introduced into the achievement testing program, and in the long run, reduces the labor needed. Furthermore, the existence of item-banks can assist in tailored testing and the use of flexi-tests (Orr, 1978; Womer, 1978). Many such tests used to monitor student progress could also be easily constructed, administered, and scored. Since sophisticated computers are common in colleges and universities, it is reasonable to expect that they can assist to an important degree in facilitating the achievement testing effort.

Communication of Measurement Results to Multiple Audiences

Even though we live in a measurement world and must daily process numerous measurement results in order to make everyday decisions, we nevertheless fail to communicate adequately the measurement results so often used in education. A classic case of poor communication exists in the case of the College Board Scholastic Aptitude Test. Even though the SAT is a household word, the fact remains that few students, parents and even faculty members understand the rationale of the test and the degree of validity its

scores have when used for various purposes.

It is not clear how one can make a massive breakthrough to solve the communication problem. Another type of norm is probably not the answer. We are already blessed with a wide variety of standard scores, and each type seems to be about as confusing to the general public as all other types.

If there is a straightforward answer to the communication problem, it probably lies in narrative reporting of test results (Ahmann and Glock, 1975, p. 217). Interesting progress has been made in this area for a number of standardized tests (for example, the Differential Aptitude Test), and it is very likely that a serious effort be made to develop suitable narrative statements and use them as the primary means of interpreting test scores, particularly scholastic aptitude scores. Obviously, properly used criterion-referenced tests will readily yield narrative reports and, should these tests be used more frequently to measure the achievement of college students, the increase in the use of narrative reports would be dramatic and conceivably would reduce some of the communication difficulties we are now experiencing. In summary, it is distressing to observe that repeatedly in higher education, the value of scores derived from well-constructed tests is virtually destroyed because we lack the proper means of communicating their meanings to students and faculty alike. Attention to this problem is long overdue in higher education.

Evaluation of Undergraduate Educational Programs

The accountability movement in higher education has grown in the 1970s and, because of declining enrollments and associated problems, will not likely continue its growth at an even greater rate in the 1980s. For this

reason if no other, it is important that systematic efforts be made to evaluate educational programs (especially the undergraduate ones) at the college and university level. Some encouraging starts have been made, such as the effort to determine the degree to which students have achieved the goals of general education by means of essay tests (Warren, 1977).

Since the primary objectives in higher education are cognitive, it stands to reason that the development of suitable achievement measures for program evaluation is a great need of the 1980s. Of course, it is too much to expect that appropriate tests can be built to measure the degree to which students have achieved all of the important goals in the many diverse educational programs found in higher education. On the other hand, would it not be better to use measures of this type to determine--at least in a rough way--the degrees of quality of various undergraduate programs, rather than to rely heavily on popularity contests based on ratings from a sample of various professional groups? As useful as some of the reports are that rank the quality of various educational programs on a national basis, they are heavily influenced by their graduate components. If a department happens to be one of the "top 20" when national rankings are reported, this no doubt means that the department has accomplished some useful things in the past at the graduate level. But what conclusions can we draw about its undergraduate programs? Probably none. The relative merits of these programs need to be evaluated separately, using highly developed achievement tests representing the major aspects of these programs.

Consistent with the foregoing, the measurement of occupational competency upon completing a baccalaureate program needs more attention in

higher education. In vocational-technical programs such efforts have been made for some time (Panitz and Olivo, 1971). Is it too much to ask that similar efforts to measure occupational competency be made in professional areas such as engineering, business administration, law, medicine, dentistry, nursing, and even teaching? Certification examinations are a step in this direction. Nevertheless, this step must be viewed as a small one.

Efforts to measure occupational competency would encompass not only paper-pencil testing, but also oral testing and performance testing. The latter would be the most complex of all, and would be very costly. Simulations such as now are being used in medicine (McGuire et al, 1976) should be thoroughly explored and applied in other areas. There is much potential in this method.

Program evaluation inevitably brings into focus the evaluation of the professor in the classroom--another fruitful area of growth for the 1980s. A useful differentiation is being made between teacher performance and teacher effectiveness. Observations made by students, peers, and administrators of a teacher's behavior when engaged in the teaching art constitute an evaluation of that teacher's performance. On the other hand, if the incremental knowledge gained by the students as a consequence of contact with a particular teacher is measured, then the teacher's effectiveness is being evaluated (Menne, 1974). In higher education, considerable attention has been given since the 1960s to the question of student rating of teacher performance. Almost no attention is given to the evaluation of teaching effectiveness. This need is a serious one and should be addressed as part of an increased emphasis on program evaluation.

It stands to reason that as better criterion-referenced and norm-

referenced achievement tests are built and used, the reasonableness of evaluating teaching effectiveness increases. Changes in levels of student behavior can be measured by the use of pretests and posttests, even though serious methodological problems exist with this approach. The changes in levels of achievement for various groups of students can be identified and used as a partial basis for examining teacher effectiveness. Such an examination requires the most careful consideration of the inputs (student characteristics, etc.) and process (all educational encounters) involved when interpreting the size of the changes in levels of achievement. In the 1980s there should be the most serious effort to date to sort out the methodological problems involved in the evaluation of teacher effectiveness, even to the point of trying to relate teacher effectiveness to teacher performance.

Systematic Assessment of the Educational Needs of College-Bound Students

Needs assessments are comparatively common when educational programs are developed in terms of federal guidelines. Such assessments, however, rarely touch higher education directly. Strange as it seems, many in higher education feel that they can accurately identify the educational needs of college-bound students without objective evidence as to the nature of these needs.

In the 1970s the National Assessment of Educational Progress systematically measured the levels of educational achievement of 17-year-olds and young adults, ages 26-35. Ten learning areas were included in the assessment, spanning from the basic skills such as mathematics and reading to the humanities and fine arts, such as art and music. These data provide for the first time profiles of the achievement of these students on a

national basis, and the changes which are occurring every three to seven years (Ahmann, 1978; 1979). Careful study of these profiles would permit inferences to be drawn with regard to the capabilities of these students as they enter either higher education programs or their life/work activities. Because the college-bound group was not identified and therefore separate results are not available, one cannot at this time delineate differences that exist with regard to the command of the basic skills and therefore the level at which it is appropriate to focus the introductory curricula in higher education, the teaching material associated with it, and the manner in which the teaching itself should be conducted. Reasonably complete data regarding the achievement patterns of college-bound youths, obtained systematically, would be of invaluable assistance in making decisions of this kind. In the meantime, we make all needed decisions on the basis of what we hope college students bring to us rather than what they actually bring to us.

Regular assessment of the educational needs of college-bound youths would provide invaluable information about how to improve the communication between secondary schools and the lower division programs in higher education. Indirectly at least, this information would also give u some ideas about the problems of integrating nontraditional students into the higher education system. Furthermore, we would have better ideas about the design of compensatory educational programs to meet special needs, should these be necessary.

At least at the lower division level, planning in higher education would be vastly improved if needs assessment were conducted systematically for college-bound youths. It must also be remembered that needs assess-

ment provides the first step of program evaluation. For example, the design of lower division programs based in part on the data yielded by needs assessment would permit the later evaluation of these programs to be cast in terms of the needs identified, and the degree to which these needs have been met. In this way, the evaluation of undergraduate academic programs can be cyclical, and hence stronger.

Summary

To be sure, there is much work to be done in the 1980s. Significant progress can be made as faculty members become more sensitive to the usefulness measurement data in program planning and evaluation in higher education. What is needed is an effort here which is comparable in quality, intensity, and duration to that which already exists in the development and use of scholastic aptitude tests for college-bound youth.

REFERENCES

Ahmann, J. S. National achievement profiles in ten learning areas. Educational Studies, 1978, 9 (in press).

Ahmann, J. S. Differential changes in levels of achievement for students in three age groups. Educational Studies, 1979, 10 (in press).

Ahmann, J. S., and Glock, M.D. Evaluating pupil growth, (5th ed.). Boston: Allyn and Bacon, 1975.

Glass, G. V. Standards and criteria. Journal of Educational Measurement, 1978, 15, 237-261.

McGuire, C. H., et al. Construction and use of written simulations. New York: The Psychological Corporation, 1976.

Menne, J. W. Teacher evaluation: Performance or effectiveness? NCME Measurement News, 1974, 17(2), 11-12.

Orr, D. B. Measurement in education in the 1980s--A federal perspective. Paper presented at the Phi Delta Kappa measurement symposium, St. Louis, 1978.

Panitz, A. and Olivo, C. T. Handbook for developing and administering occupational competency testing, (Volume 3). New Brunswick, N.J.: Rutgers, The State University, 1971.

Shoemaker, D. M. Toward a framework for achievement testing. Review of Educational Research, 1975, 45, 127-147.

Trow, W. C. Grades and objectives in higher education. Educational Record, 1968, 49, 85-91.

Warren, J. R. Academic competences in general education. Princeton, N.J.: Educational Testing Service, 1977.

Womer, F. B. Measurement in education for the 1980s. Paper presented at the Phi Delta Kappa measurement symposium, St. Louis, 1978.